KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Sovereign credit rating determinants: the impact of the European debt crisis

Reusens P, Croux C.

KBI_1615

Sovereign credit rating determinants:
the impact of the European debt crisis

Peter Reusens[a,*], Christophe Croux[a]

[a]*Faculty of Economics and Business, KU Leuven, Belgium.*

Abstract

This paper compares the importance of different sovereign credit rating determinants over time, using a sample of 90 countries for the years 2002-2015. Applying the composite marginal likelihood approach, we estimate a multi-year ordered probit model for each of the three major credit rating agencies. After the start of the European debt crisis in 2009, the importance of the financial balance, the economic development and the external debt increased substantially and the effect of eurozone membership switched from positive to negative. In addition, GDP growth gained a lot of importance for highly indebted sovereigns and government debt became much more important for countries with a low GDP growth rate. These findings provide empirical evidence that the credit rating agencies changed their sovereign credit rating assessment after the start of the European debt crisis.

Keywords: Composite marginal likelihood, Credit rating agencies, European debt crisis, Multi-year ordered probit model, Sovereign credit rating determinants

JEL: C33, C35, F34, G24, H63

1. Introduction

A sovereign credit rating is a measure of the creditworthiness of a sovereign government assigned by a credit rating agency (CRA). Each sovereign credit rating is determined by a rating committee, which assesses the different factors that drive the sovereign's creditworthiness. Rather than computing a fixed weighted average of these factors, CRAs can vary the relative importance of the various factors over time, in response to changing macroeconomic circumstances (Kiff *et al.*, 2010). For instance, Fitch (2014) states they attach more importance to the sovereign public finance ratios and financing

*Corresponding author: Faculty of Economics and Business, KU Leuven, Naamsestraat 69 B-3000 Leuven Belgium, Tel: +32 16 326728.

Email addresses: peter.reusens@kuleuven.be (Peter Reusens), christophe.croux@kuleuven.be (Christophe Croux)

flexibility during crisis periods and Gaillard (2012) argues that, before the outbreak of the European debt crisis, CRAs attached too much value to both the advanced economy status and eurozone membership of Greece. Even though the CRAs regularly publish reports in which they identify the different ingredients of the sovereign credit rating, further judgmental adjustments are made by the rating committee.[1] Therefore, the actual degree of importance of the different variables and their change over time is not known. In this paper, we quantify, for the three major rating agencies Standard and Poor's (S&P), Moody's and Fitch, how the importance of different sovereign credit rating determinants changed after the start of the European debt crisis.

Starting with Cantor and Packer (1996), an empirical literature has emerged that analyzes the importance of the determinants of sovereign credit ratings using historical data. In their seminal paper, Cantor and Packer (1996) report that their single-year linear regression model with eight macroeconomic variables could explain more than 90% of the variability of the sovereign credit ratings for 1995. In particular, they find a statistically significant effect of the variables GDP per capita, GDP growth, inflation, external debt, the economic development and default history. Subsequent research has confirmed the importance of these variables in explaining the sovereign credit rating (Afonso *et al.*, 2011; Gaillard, 2012; Gartner *et al.*, 2011).

Only a few papers have compared the importance of the different credit rating determinants over time. These papers predominantly analyze the change after the 1997-1998 Asian crisis and mostly use *linear* regression models. Monfort and Mulder (2000) compare estimated coefficients of their panel linear regression model between subperiods 1994-1995, 1996-1997 and 1997-1998. They find stable coefficients across subperiods, with the exception of the export growth rate. Bissoondoyal-Bheenick (2005) estimate single-year ordered probit regression models for the years 1995 to 1999 and finds that mostly the same variables are statistically significant over the different years. Finally, Afonso *et al.* (2007), estimating a panel linear regression model separately for the period 1996-2000 and 2001-

[1]Standard and Poor's scores five key factors of a sovereign's degree of creditworthiness on a six point scale. While the calibration of both the scoring procedure of each factor and the procedure of combining the scores into a single credit rating has been more objectively documented as from 2011, qualitative judgment still remains important in this rating process (S&P, 2014b). Also Moody's (2015b) uses a scorecard which maps different indicators to four key factors, which are then combined to an initial sovereign credit rating. Although Moody's provides indicative weights of the different determinants of each of these factors, they emphasize that the actual weights can substantially deviate because of supplementary adjustments based on qualitative judgment. Finally, the rating process of Fitch (2014) starts from the rating prediction of a linear regression model in which 19 variables are regressed on historical Fitch sovereign credit ratings and which is yearly re-estimated for a sample starting in 2000. Also here, the rating committee makes substantial changes to this initial rating prediction.

2006, conclude that most estimated coefficients are similar across subperiods, which they interpret as evidence for a rather stable credit rating process over time.

This paper builds on above literature that compares the importance of the different factors of sovereign credit ratings over time. Using a sample of 90 countries for the period 2002-2015, we investigate if and how the importance of the sovereign credit rating determinants changed after the start of the European debt crisis in 2009. This analysis is performed for each of the three major rating agencies Standard and Poor's (S&P), Moody's and Fitch and the focus is predominantly on common patterns over time. Estimating a multi-year ordered probit model using a composite marginal likelihood estimation approach, we are the first to take into account both the ordinal nature of the sovereign credit rating and the serial correlation of the error terms. We compare the *importance* of the different credit rating determinants over time, whereas the existing literature that uses an ordered probit model, has only analyzed the statistical significance and the sign of the estimated coefficients. A difficulty is that the coefficients of different ordered probit models are not directly comparable over time, because their scaling depends on the unobserved degree of residual variation (Allison, 1999).

While previous literature has predominantly focused on the impact of the Asian crisis on the importance of the different credit rating determinants, we analyze the impact of the European debt crisis. For each of the three major credit rating agencies, we find that, after the start of the European debt crisis in 2009, the importance of the financial balance, the economic development and the external debt increased substantially and that the effect of eurozone membership switched from positive to negative. In addition, GDP growth gained a lot of importance, especially for highly indebted sovereigns, and government debt became much more important, especially for countries with a low GDP growth rate. These findings provide empirical evidence that the CRAs changed their sovereign credit rating assessment after the start of the European debt crisis.

Our paper is organized as follows. Section 2 discusses the data and Section 3 presents the multi-year ordered probit model. Then, Section 4 discusses the results and Section 5 concludes our findings.

2. Data

We use data for the $T = 14$ years between 2002 and 2015. For the three major rating agencies S&P, Moody's and Fitch, we have a balanced panel dataset for respectively 85, 90 and 69 advanced and emerging countries, listed in Table A.2 of Appendix A. End of year sovereign credit ratings are obtained from S&P, Moody's and Fitch and the different rating categories are shown in Table A.1 of Appendix A.

Table 1: Definitions of the explanatory variables, the source of the data and the expected sign of the impact on the credit rating.

Variable name	**Definition**	**Source**	**Sign**
GDP per capita	GDP per capita, PPP (international dollars)	IMF, WEO Oct2015	+
Government debt	General government gross debt (% of GDP)	IMF, WEO Oct 2015	-
GDP growth	Real GDP growth (annual %)	IMF, WEO Oct 2015	+
Eurozone membership	Member country of the European Monetary Union	ECB	+/-
Financial Balance	Financial balance (% of GDP)	Moody's (2015a)	-
Economic development	Member country of the OECD	OECD	+
External debt	External debt (% of GDP) (developing countries)	Moody's (2015a)	-
Current account	Current account balance (% of GDP)	IMF, WEO Oct 2015	+/-
Inflation	Inflation, end of period consumer prices (annual %)	IMF, WEO Oct 2015	-
Default history	Sovereign default since 1975	Beers and Nadeau (2015)	-

We use ten rating determinants in our model. We include GDP per capita, government debt, GDP growth, inflation, financial balance, external debt, current account and dummy variables for economic development and default history, which have been previously shown to be important drivers for the creditworthiness (Afonso *et al.*, 2011; Cantor and Packer, 1996; Elkhoury, 2007; Gaillard, 2012). In addition, we include the dummy variable for eurozone membership, which importance is expected to have changed after the European debt crisis. The data definitions, data sources and expected sign of the effect of the determinants on the credit rating are shown in Table 1 and summary statistics are shown in Table 2.

Table 2: Summary statistics for each variable, computed over all observations that have an S&P rating for the years 2002 to 2015.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
GDP per capita	1353.00	10300.00	20110.00	24940.00	35160.00	149600.00
Government debt	0.06	28.50	43.46	52.11	67.63	246.20
GDP growth	-15.14	1.55	3.30	3.44	5.47	26.17
Eurozone membership	0.00	0.00	0.00	0.17	0.00	1.00
Financial balance	-32.30	-4.38	-2.30	-1.86	-0.10	40.80
Economic development	0.00	0.00	0.00	0.36	1.00	1.00
External debt	0.00	0.00	25.80	33.38	45.95	965.00
Current account	-53.56	-4.55	-0.83	0.04	3.67	45.22
Inflation	-4.90	1.61	3.08	4.64	5.86	190.00
Default history	0.00	0.00	1.00	0.52	1.00	1.00

For each of the determinant, the expected sign of its effect on the credit rating is motivated as follows:

- *GDP per capita*: Countries with a higher GDP per capita are expected to have a higher sovereign credit rating, conditioning on the other variables in the model. These countries have a higher potential tax base and they often have a sound political and institutional stability.

- *Government debt*: Countries with a higher level of government debt relative to GDP are expected to have a lower sovereign credit rating.

- *GDP growth*: Countries with a higher GDP growth rate are expected to have a higher sovereign credit rating, because a higher GDP growth rate is indicative for a higher future GDP growth rate, which increases the future potential tax base and reduces the future government debt to GDP ratio.

- *Eurozone membership*: Membership to the eurozone monetary union[2] has an ambiguous impact on the sovereign credit rating of its member states. On the one hand, enforceable rules for fiscal discipline, such as the Stability and Growth Pact, increase the fiscal credibility of its member states (Afonso *et al.*, 2011; Gartner *et al.*, 2011). Eurozone membership also provides several economic advantages for member states, such as decreased transaction costs and reduced price uncertainty, which lead to increased trade and economic activity. Another advantage is that the euro is an actively traded currency, such that the member country can more easily issue debt in domestic currency (S&P, 2014a). On the other hand, many researchers such as De Grauwe and Ji (2013), state that member countries in a monetary union are prone to a self-fulfilling liquidity crisis. As these member countries cannot force the central bank to alleviate a liquidity crisis by buying their government debt, they can face higher interest rates during a liquidity crisis. This high interest rate, together with the fact that economic growth cannot be boosted through currency depreciation, implies that a liquidity crisis can easily spillover into a solvency crisis.

- *Financial balance*: A positive financial balance relative to GDP signals that the government is able and willing to increase taxes or reduce expenses in order to service its debt.

- *Economic development*: Countries that are classified as economically developed, are expected to have a higher credit rating. They are perceived to have attained a certain threshold of economic development for which default is very unlikely. In addition, these countries are often strongly

[2]Note that we do not investigate the effect of membership to other currency unions, since too few of such members received a sovereign credit rating.

integrated with the world economy, such that a default is less likely, as foreign creditors can more easily disrupt trade or seize assets abroad in case of default (Cantor and Packer, 1996).

- *External debt*: Countries with a high external debt relative to GDP have a high total debt burden, such that additional taxes or reduced government expenses are needed in order to reduce the government's debt or to support over-indebted domestic borrowers (Afonso *et al.*, 2011). Given that data on the external debt is missing for many industrialized countries, we only analyze its effect for developing countries (as defined in Moody's, 2015a), by setting the external debt to zero for the industrialized countries, in line with Hill *et al.* (2010) and Afonso *et al.* (2011).

- *Current account*: The current account balance of a country has an unclear impact on its sovereign credit rating. While a current account surplus is expected to positively impact the credit rating, the effect of a current account deficit on the credit rating depends on the productivity of the investment it finances.

- *Inflation*: A high inflation rate may be a symptom of macroeconomic problems and can lead to dissatisfied inhabitants and corresponding political instabilities (Afonso *et al.*, 2011; Bissoondoyal-Bheenick *et al.*, 2006). This negative effect is partly offset because a high inflation also lowers the real stock of outstanding government domestic currency debt and because an inflation that is too low may lead to a deflationary spiral.

- *Default history*: Sovereigns that have previously defaulted on their debt, are seen as being less willing to repay their debt.

Finally, we also add the interaction term[3] between GDP growth and government debt to the model, because GDP growth matters more for the sovereign's creditworthiness, if the level of government debt is high. In particular, an increase in the GDP growth rate, reduces the future government debt ratio, hence increases the sovereign's creditworthiness, by an amount proportional to the debt ratio.[4] We expect this interaction effect to be positive.

[3]The interaction term is computed as the product of the centered GDP growth rate and the centered government debt ratio, in which the overall mean is used to center the variables (i.e. 3.348 for GDP growth rate and 52.455 for government debt).

[4]Indeed, keeping the total real amount of government debt constant, we would have that next year's government debt ratio equals $\frac{d}{1+g} \approx d - dg$, with g the real GDP growth rate and d the present debt ratio.

3. Methodology

In early research on the determinants of sovereign credit ratings, a linear regression model was used in which the dependent variable credit rating was transformed to a linear scale. We believe this linear model to be inappropriate for two reasons. First, the linear regression model assumes that the absolute distances in the underlying degree of creditworthiness between subsequent credit rating categories are equally spaced. This assumption is not realistic for credit ratings as they are only ordinal measures for the sovereign's degree of creditworthiness, see e.g. Afonso *et al.* (2011), Bissoondoyal-Bheenick (2005) and Mora (2006). Second, McKelvey and Zavoina (1975) have shown that, even if the degree of creditworthiness were equally spaced between rating categories, applying linear regression to ordinal data would still result in a bias in the estimated coefficients. Christensen (2015) state that this bias of the linear model is small only if there are many response categories and the responses do not pile up in the end categories. Given that 17%, 19% and 21% percent of ratings has the highest rating category for S&P, Moody's and Fitch respectively, the bias is hence expected to be considerable.

The ordered regression model is not subject to the above discussed disadvantages of the linear regression model and it is increasingly used for modeling sovereign credit ratings. Bissoondoyal-Bheenick (2005) and Gaillard (2012) use a single-year ordered regression model for sovereign credit ratings. Also, Hill *et al.* (2010) and Hu *et al.* (2002) estimate an ordered regression model, pooling data from multiple years. These single-year and pooled ordered probit models do not exploit the panel data structure of sovereign credit ratings, collected over a span of fourteen years. Subsequently, Afonso *et al.* (2009) and Mora (2006) estimate a panel ordered probit model, respectively using random and fixed effects. However, as these models assume that both the regression coefficients and threshold parameters are constant over time, they do not allow for a comparison of the coefficients over time.

We use a multi-year ordered probit regression model, which allows for time variation in the regression coefficients and explicitly models the correlation between the error terms over the years. Our model is similar to the cross-sectional multivariate ordered response model used by Bhat *et al.* (2010) and Ferdous *et al.* (2010) to assess the determinants of the level of non-work activities for different activity types.

3.1. The multi-year ordered probit regression model

Consider the latent regression equation

$$Y_{it}^{*} = \beta_t' x_{it} + \nu_{it} \tag{1}$$

for i in $1,...,N$ and t in $1,...,T$, where N is the number of countries and T is the number of time periods, Y^*_{it} is an unobserved latent variable measuring the degree of creditworthiness of sovereign i at time t, x_{it} is a vector of p explanatory variables of sovereign i at time t, β_t is a vector of unknown parameters at time t, and $(\nu_{i1},...,\nu_{iT})$ are jointly standard normally distributed error terms with correlation matrix Σ.[5] In order to reduce the number of free parameters in the correlation matrix Σ, we hypothesize, in line with Varin and Czado (2010), that the error term of each sovereign i follows an autoregressive process of order one with common autoregressive parameter ρ, so that the element of Σ at row s and column t, is given by $\Sigma_{st} = \rho^{|t-s|}$.

The threshold specification is given by

$$Y_{it} = \begin{cases} 1 & \text{if } -\infty < Y^*_{it} < \tau^1_t \\ l & \text{if } \tau^{l-1}_t \leq Y^*_{it} < \tau^l_t \text{ for } l = 2, \ldots, C_t - 1 \\ C_t & \text{if } \tau^{C_t-1}_t \leq Y^*_{it} < \infty \end{cases} \tag{2}$$

for i in $1,...,N$ and t in $1,...,T$, where Y_{it} is the observed credit rating, τ^l_t is a threshold parameter and C_t represents the number of observed rating categories in the sample for time t.[6] For notation purpose, we label $\tau^0_t = -\infty$ and $\tau^{C_t}_t = \infty$. In sum, the parameters of the model are the pT coefficients β_t, the $\sum_{t=1}^{T}(C_t - 1)$ threshold parameters τ^l_t and the correlation parameter ρ, and we collect them in the vector θ.

3.2. The likelihood function

The likelihood function is given by

$$L(\theta) = \prod_{i=1}^{N} L_i(\theta), \tag{3}$$

where $L_i(\theta)$ is the likelihood for sovereign i, given by

$$\begin{aligned} L_i(\theta) &= P\left(Y_{i1} = y_{i1}\ , \ldots,\ Y_{iT} = y_{iT}\right) \\ &= \int\limits_{\nu_{i1}=\tau_1^{y_{i1}-1}-\beta_1' x_{i1}}^{\tau_1^{y_{i1}}-\beta_1' x_{i1}} \cdots \int\limits_{\nu_{iT}=\tau_T^{y_{iT}-1}-\beta_T' x_{iT}}^{\tau_T^{y_{iT}}-\beta_T' x_{iT}} \phi(\nu_{i1},...,\nu_{iT};\Sigma)\, d\nu_{i1}...d\nu_{iT}, \end{aligned} \tag{4}$$

[5]The scaling of the variances of the error terms ν_{it} to 1 and the absence of intercept coefficients are necessary to identify the model parameters.

[6]If a certain rating category is not observed in the sample, there is no information in the data to identify its corresponding threshold parameter. Although each CRA has 21 rating categories, the number of *observed* different rating categories C_t varies over the years between 16 and 19 for S&P, between 17 to 19 for Moody's and between 15 to 18 for Fitch.

where y_{it} is the *observed* category number of variable Y_{it} and $\phi(\nu_{i1}, ..., \nu_{iT}; \Sigma)$ is the density of the multivariate normal distribution with mean *zero* and correlation matrix Σ.

Since the T-dimensional integral in (4) cannot be easily computed for dimensions larger than two, a classical maximum likelihood estimation is not feasible. One could approximate the T-dimensional integral in (4) using simulation techniques, but the corresponding simulated maximum likelihood estimator should not be used for high dimensional multivariate ordered response settings, due to computational convergence issues (Bhat *et al.*, 2010).

3.3. The composite likelihood estimator

The composite likelihood estimator $\hat{\theta}$ (Bhat *et al.*, 2010) maximizes the composite likelihood function

$$L^C(\theta) = \prod_{i=1}^{N} L_i^C(\theta), \tag{5}$$

with $L_i^C(\theta)$ the pairwise marginal likelihood function for sovereign i

$$L_i^C(\theta) = \prod_{s=1}^{T-1} \prod_{t=s+1}^{T} P\left(Y_{is} = y_{is}, Y_{it} = y_{it}\right), \tag{6}$$

where y_{is} and y_{it} denote the *observed* category of variables Y_{is} and Y_{it}, and $P\left(Y_{is} = y_{is}, Y_{it} = y_{it}\right)$ is the probability of their joint occurrence. It is a consistent and asymptotically normally distributed estimator with covariance matrix $Cov(\hat{\theta})$, and it is only slightly less efficient than the full maximum likelihood estimator. Complete expressions for the composite likelihood function, the estimates of $Cov(\hat{\theta})$ and implementation details of the composite likelihood estimator are given in Appendix B.

3.4. Comparing coefficients over time in the multi-year ordered probit model

The estimated coefficient $\hat{\beta}_t^v$ represents the estimated effect for time t of a one unit increase in the variable v on the underlying degree of creditworthiness Y_{it}^*, keeping the other variables constant. However, a direct comparison over time of these estimated coefficients is not meaningful, because the unit of measurement of the unobserved underlying degree of creditworthiness Y_{it}^* differs over time (Allison, 1999). This change in unit of measurement arises because the variances of the error terms in the ordered regression model are scaled to one.

As a solution, we apply the approach of Hoetker (2004) and Hoetker (2007), originally proposed for comparing coefficients across binary choice models. This entails a scaling of the coefficients across time. Let GDP per capita be the first variable. We will analyze the ratio R_t^v, which we call the *importance* of the variable,

$$R_t^v = \frac{\beta_t^v}{\beta_t^{GDP}} \text{ for } v \text{ in } 2, ..., p \text{ and } t \text{ in } 1, ..., T, \tag{7}$$

where β_t^{GDP} is the coefficient of the variable GDP per capita. The interpretation of this ratio is that, ceteris paribus, a one unit increase in the variable of interest is expected to have the same effect on the degree of creditworthiness as an increase in GDP per capita by the amount equal to the value of this ratio. This ratio can also been interpreted in terms of the 'compensating variation' used in Boes and Winkelmann (2006), Train (1998) and Train (2003): it represents the required increase in GDP per capita necessary to offset a one unit decrease of the variable v, such that the sovereign's degree of creditworthiness remains the same.

The importance R_t^v is estimated by the sample counterpart of (7), where the coefficients β_t^v and β_t^{GDP} are replaced by their composite likelihood estimate of Section 3.3. The estimated covariance matrix of $\hat{R}_t^v$ is obtained using the Delta method and the estimated covariance matrix, given in Appendix B.2.

4. Results

4.1. Estimated importance of the credit rating determinants

For the different determinants v and the different time periods t, Figures 1, 2 and 3 show the estimated ratio $\hat{R}_t^v$, as defined in Section 3.4, for S&P, Moody's and Fitch, respectively.[7] As elaborated in Section 3.4, this ratio quantifies the importance of each determinant, as it represents the required increase in GDP per capita necessary to offset a one unit decrease of the determinant such that the sovereign's degree of creditworthiness remains the same. The figures also show the 95% confidence bounds. We detect important changes in the importance of the different variables after the start of the European debt crisis in 2009. Averages for the estimated importance $\hat{R}_t^v$ over the period 2002-2008 and over the period 2009-2015 are shown in Table 3.

[7] Note that for each variable v and time period t, the estimated importance $\hat{R}_t^v$ and the estimated coefficient $\hat{\beta}_t^v$ have the same sign and a similar significance pattern, because the estimated ordered probit coefficient of GDP per capita $\hat{\beta}_t^{GDP}$ is positive and significant for each year (the estimated ordered probit coefficients are available upon request).

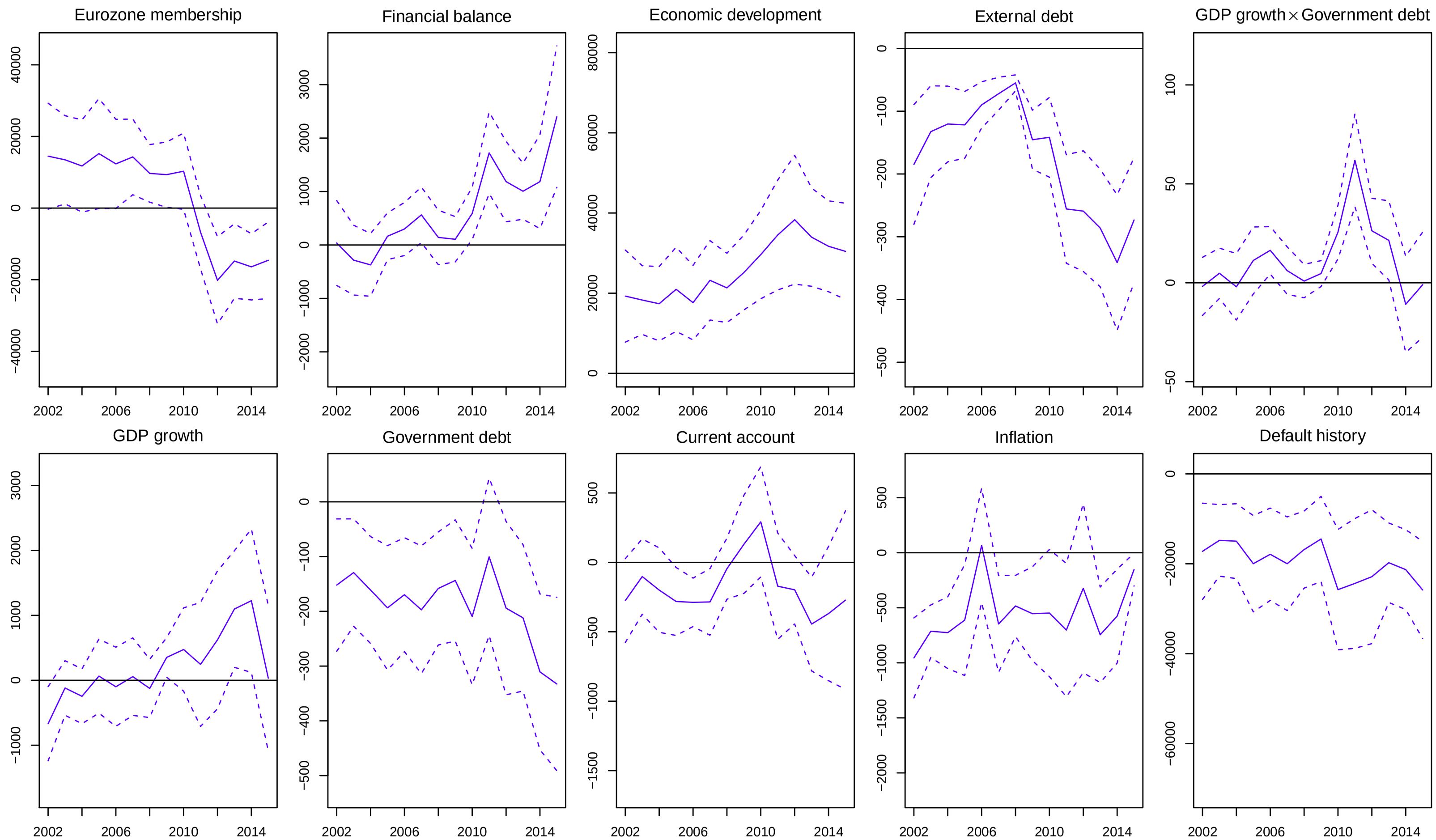

Figure 1: Estimated importance of each variable for S&P as a function of time. The dashed lines are pointwise 95% confidence bounds.

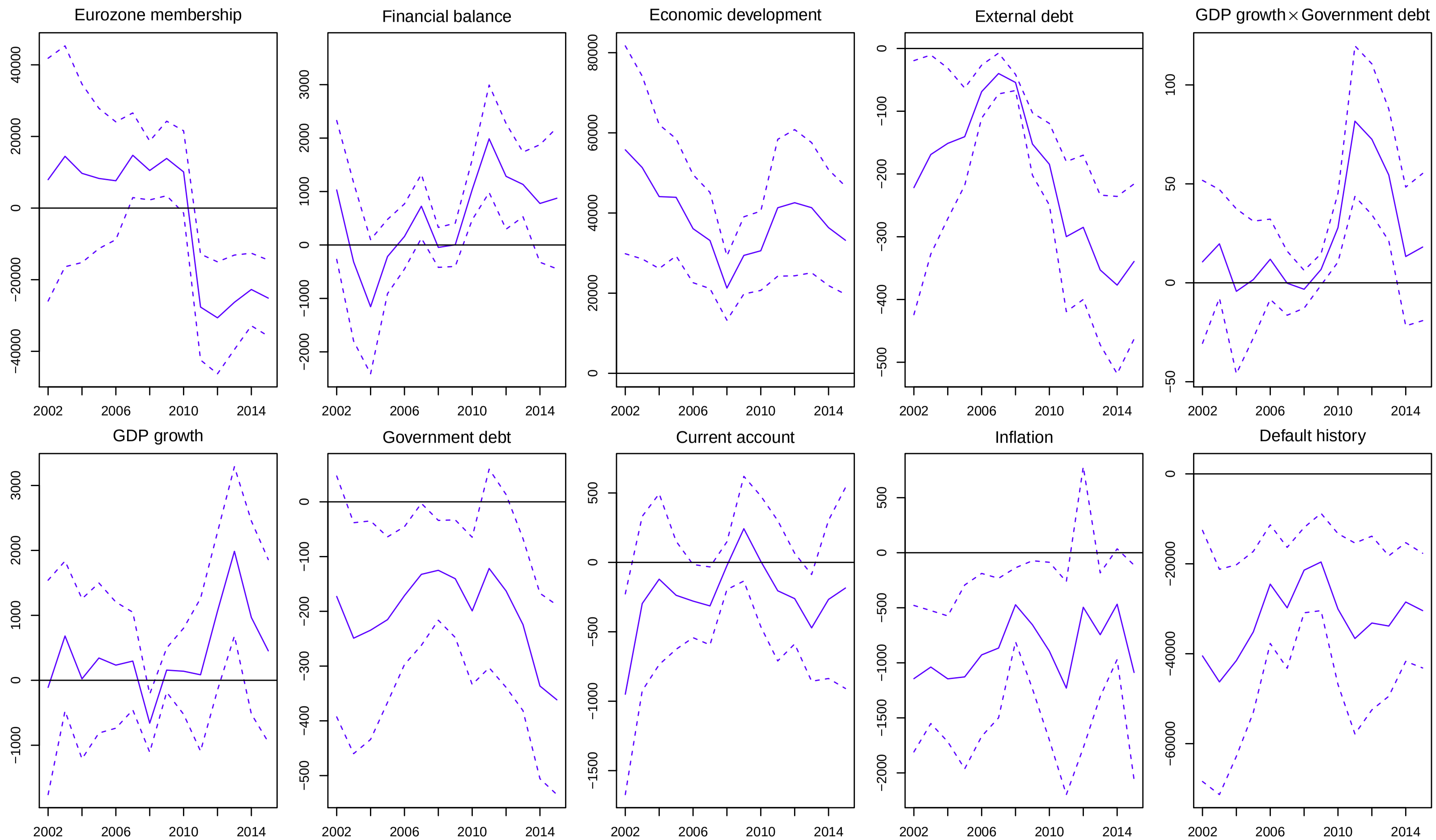

Figure 2: Estimated importance of each variable for Moody's as a function of time. The dashed lines are pointwise 95% confidence bounds.

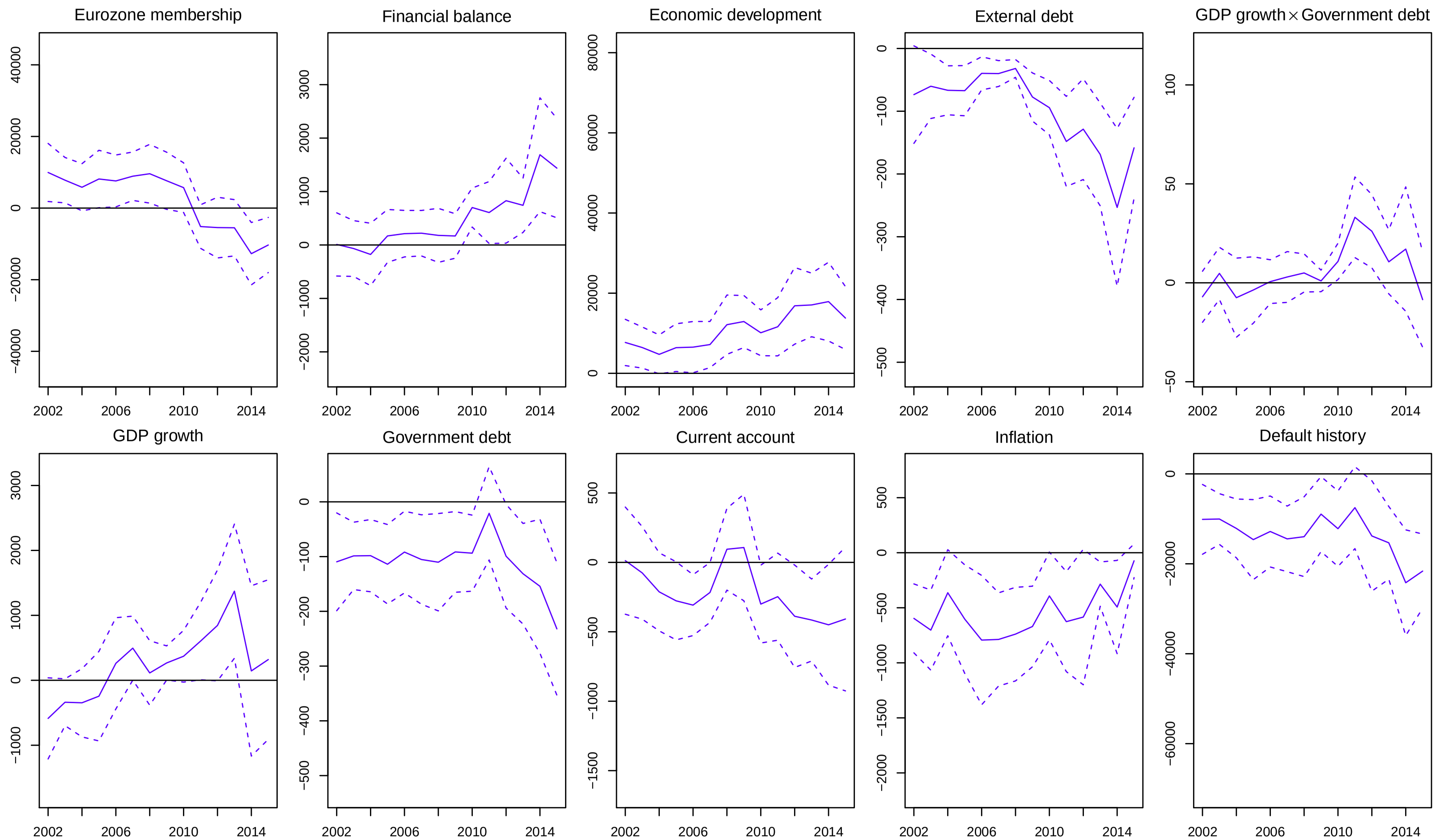

Figure 3: Estimated importance of each variable for Fitch as a function of time. The dashed lines are pointwise 95% confidence bounds.

Table 3: For each variable and each CRA, the table shows the average estimated importance $\hat{R}_t^v$ for the period 2002-2008 (left panel) and for the period 2009-2015 (middle panel), as well as the P-values of the Wald hypothesis test that the average importance is the same for both periods (right panel).

	Average $\hat{R}^v_{2002-2008}$			**Average $\hat{R}^v_{2009-2015}$**			**H_0: No break in 2009**		
	S&P	Moody's	Fitch	S&P	Moody's	Fitch	S&P	Moody's	Fitch
Eurozone membership	13038	10464	8249	-7581	-15502	-3688	**0.000**	**0.002**	**0.000**
Financial balance	78	26	78	1171	1013	882	**0.001**	**0.031**	**0.004**
Economic development	19720	40804	7303	31961	36393	14321	**0.006**	0.552	**0.004**
External debt	-111	-121	-54	-243	-284	-147	**0.000**	**0.001**	**0.000**
GDP growth × Government debt	5	5	-1	18	39	13	0.072	**0.008**	0.054
GDP growth	-163	115	-92	578	693	560	**0.030**	0.240	**0.037**
Government debt	-166	-186	-104	-215	-221	-118	0.206	0.564	0.648
Current account	-211	-317	-140	-148	-163	-300	0.655	0.478	0.274
Inflation	-581	-960	-655	-514	-796	-446	0.706	0.632	0.179
Default history	-17379	-34163	-12593	-22046	-30312	-14803	0.219	0.591	0.475

The effect of eurozone membership, fiscal balance, economic development and external debt on the credit rating changed substantially after 2009. *(i)* While the estimated importance of eurozone membership was statistically significant and positive before 2009, on average about 10000, it substantially decreased after 2009 and became significant and negative, on average about -8000, -15000 and -4000 for S&P, Moody's and Fitch, respectively. *(ii)* While before 2009, the importance of the financial balance to GDP ratio was insignificant, it became significant and positive afterwards. After 2009, a one percentage point increase in the financial balance is expected to have the same effect on the credit rating as an increase in GDP per capita by about 1000$, on average. *(iii)* The estimated importance of economic development increased after 2009 for S&P and Fitch. For Moody's, however, the picture is less clear. *(iv)* The importance of external debt is significant and negative for all years, but it decreased from about -100 before 2009 to about -250 for the period after 2009.

Also the effect of government debt and GDP growth rate on the credit rating changed substantially after 2009. The graphs of the importance of GDP growth and government debt correspond to a country with an average value for these variables. *(i)* The interaction term between GDP growth rate and government debt is significant and positive between 2009 and 2013. *(ii)* For a country with an average debt ratio, the importance of GDP growth rate was insignificant for the years before 2009 and positive and often significant for the years after 2009. For a highly indebted sovereign with a government debt ratio of 100% (i.e. the 90% percentile of the government debt ratio in our sample), the total effect of a one percentage point increase of GDP growth after 2009 is equivalent to an increase in GDP per capita of about 1400$, 2600$ and 1200$ for S&P, Moody's and Fitch, respectively. In contrast, for a lowly indebted sovereign with a government debt ratio of 20% (i.e. the 10% percentile of the government debt ratio), the total effect of a one percentage point increase of GDP growth has remained close to zero. *(iii)* For a country with an average GDP growth rate, the importance of government debt has the expected negative sign and is significant for most years; it increased in magnitude by about 20% after 2009. For countries with a GDP growth rate of -1% (the 10% percentile), the importance of government debt increased substantially in magnitude after 2009 to -300, -400, -180 for S&P, Moody's and Fitch, respectively, whereas the importance of government debt remained equal to about -150 for countries with a growth rate of 6% (the 90% percentile).

The estimated importance of the other variables remained relatively constant over the sample period. *(i)* The current account balance is insignificant for nearly all years. *(ii)* The estimated effect of a one percentage point decrease in inflation is significant for most years and corresponds to an increase in GDP per capita of about 550\$, 880\$ and 550\$ for S&P, Moody's and Fitch, respectively. *(iii)* The estimated importance of the default history is negative and significant. In particular, the impact of having defaulted in the last decades is equivalent to an increase in GDP per capita by about -19000\$, -32000\$ and -14000\$ for S&P, Moody's and Fitch, respectively.

Finally, the estimated autoregressive parameter ρ of the error term ν_{it} is large: 0.965, 0.953 and 0.961 for S&P, Moody's and Fitch, respectively. In words, a sovereign which received a higher (lower) rating than expected based on the rating determinants for a given year, is also very likely to have a higher (lower) rating than expected for the following years. This high persistence of the error terms strengthens the benefit of using a multi-year ordered probit model as the efficiency gain over the estimation of a single-year ordered probit model is substantial when the correlation of the error terms is large. Note that we did not discuss the time variation of the threshold parameters, given that their interpretation is not of interest for this paper (results are available upon request).

4.2. Test for a break

We perform two hypothesis tests. First, we test for each variable v, the hypothesis that its importance is constant across all years

$$H_0 : R_1^v = ... = R_T^v. \tag{8}$$

Table 4 shows the P-value of the Wald test for this null hypothesis. For most variables and CRAs, we reject this null hypothesis at the 5% significance level, which motivates the use of a model that allows for time variation in the ordered probit coefficients, rather than a fixed coefficients panel model.

Table 4: P-values of the Wald test for the null hypothesis that the importance R_t^v is equal for all years.

	H_0: Equality for all years		
	S&P	Moody's	Fitch
Eurozone membership	**0.000**	**0.000**	**0.006**
Financial balance	**0.001**	**0.000**	0.079
Economic development	0.182	**0.001**	**0.038**
External debt	**0.000**	**0.000**	**0.004**
GDP growth × Government debt	**0.000**	**0.018**	**0.018**
GDP growth	**0.005**	**0.013**	**0.036**
Government debt	**0.035**	**0.022**	**0.038**
Current account	0.120	0.059	0.091
Inflation	**0.000**	0.238	**0.000**
Default history	0.132	**0.039**	**0.004**

Second, we test, for each variable v in $2, ..., p$, the hypothesis that the average importance is equal before and after 2009, which is the start of the European debt crisis

$$H_0 : \frac{1}{7} \sum_{t=2002}^{2008} R_t^v = \frac{1}{7} \sum_{t=2009}^{2015} R_t^v. \tag{9}$$

The right panel of Table 3 shows the P-value of the Wald test for this null hypothesis. For most CRAs, the hypothesis of no break in 2009 is strongly rejected for eurozone membership, the financial balance, the economic development, the external debt, GDP growth and the interaction effect between GDP growth and government debt. Therefore, the previously discussed changes in the importance of these variables after 2009 are also statistically significant.

4.3. Discussion

In sum, for S&P, Moody's and Fitch, we find that the importance of the financial balance, the economic development and the external debt increased substantially in magnitude after 2009 and that the effect of eurozone membership switched from positive to negative. In addition, GDP growth and government debt, as well as their interaction, gained much importance, such that the positive effect of GDP growth on the credit rating became considerable, especially for highly indebted sovereigns, and that the negative effect of government debt became large, especially for low growth countries. These empirical findings indicate a change in the sovereign credit rating assessment of CRAs after the start of the European debt crisis. There are several possible explanations for this change.

A first explanation is that credit rating agencies had badly judged the importance of the different credit rating determinants with respect to default risk before 2009 and that they have permanently adjusted their rating methodology after the European debt crisis experience. Kiff *et al.* (2010) provide a similar argument for the change of importance of short term debt after the Asian crisis.

A second explanation is that this change only holds temporary for the duration of the European sovereign debt crisis. This interpretation would be in line with Fitch (2014), who states that during crisis periods, a higher weight is attached to sovereign's finance ratios (as government debt and financial balance) and financing flexibility. The larger weight of the sovereign's financing flexibility is reflected in the negative effect of eurozone membership after 2009, since eurozone member countries cannot force the central bank to provide them with sufficient liquidity.

4.4. Model fit

We compare the model fit of our multi-year ordered probit model to that of the single-year ordered probit model, the pooled ordered probit model, the multi-year seemingly unrelated linear regression (SUR) model and the single year OLS linear regression model.[8] Table 5 shows the mean absolute error (MAE)[9] and, in line with Afonso *et al.* (2007), the percentage of prediction errors that are within x notches, where x ranges between 0 and 6. The single-year and multi-year ordered probit outperform the OLS and SUR linear regression and the pooled ordered probit. They have on the whole the lowest MAE (about 1.5) and the highest proportion of the ratings are correctly predicted within 0 and 1 notches (about 30% and 55%).

Table 6 presents the frequency of upgrades and downgrades of the actual and the predicted ratings, together with the percentage of correctly predicted up/down grades, i.e. the percentage of time points where the sign of the change of the actual and predicted ratings coincide (similar as in Afonso *et al.*, 2007). The multi-year ordered probit model performs best: averaged over the three CRAs, it correctly predicts 44% of the rating upgrades and 56% of the rating downgrades. Finally, note that rating changes occur more often for the fitted ratings of all models (on average 23% for upgrades and 19% for downgrades) compared to actual ratings (on average 14% for upgrades and

[8]For the OLS and SUR linear regression models, we have transformed the 21 credit rating categories of Table A.1 to an equally spaced linear scale ranging between 1 and 21, in line with Bissoondoyal-Bheenick *et al.* (2006) and Giacomino (2013). In line with Afonso *et al.* (2007), we round the predicted value of the OLS and SUR model to the nearest integer between 1 and 21. Finally, we compute the prediction of the multi- and single-year ordered probit models $\hat{Y}_{it}$, as the rating category l for which $\hat{\tau}_t^{l-1} \leq \hat{\beta}_t' x_{it} < \hat{\tau}_t^{l}$.

[9]For the calculation of the mean absolute error, we consider the distance between subsequent rating categories to be one, in line with the linear scale used for the OLS and SUR linear regression models.

Table 5: Mean absolute error (MAE) and the percentage of ratings that are correctly predicted within x notches, where x ranges between 0 and 6.

		MAE	% correctly predicted within x notches						
			$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$	$x = 5$	$x = 6$
S&P	Multi-year ordered probit	1.57	0.28	0.56	0.77	0.89	0.96	0.99	0.99
	Single-year ordered probit	1.51	0.31	0.58	0.78	0.90	0.95	0.98	0.99
	Pooled ordered probit	1.72	0.26	0.54	0.74	0.86	0.94	0.97	0.99
	SUR linear regression	1.79	0.18	0.49	0.74	0.88	0.96	0.98	0.99
	OLS linear regression	1.60	0.22	0.55	0.79	0.90	0.96	0.99	0.99
Moody's	Multi-year ordered probit	1.59	0.30	0.55	0.77	0.88	0.95	0.98	0.99
	Single-year ordered probit	1.53	0.32	0.57	0.78	0.89	0.95	0.97	0.99
	Pooled ordered probit	1.81	0.26	0.50	0.72	0.85	0.93	0.96	0.98
	SUR linear regression	1.88	0.17	0.48	0.71	0.87	0.94	0.97	0.99
	OLS linear regression	1.62	0.22	0.56	0.77	0.90	0.96	0.98	1.00
Fitch	Multi-year ordered probit	1.28	0.39	0.67	0.84	0.93	0.96	0.98	0.99
	Single-year ordered probit	1.24	0.39	0.67	0.87	0.94	0.96	0.98	0.99
	Pooled ordered probit	1.52	0.31	0.61	0.80	0.90	0.94	0.97	0.98
	SUR linear regression	1.55	0.25	0.57	0.81	0.92	0.97	0.98	0.99
	OLS linear regression	1.37	0.25	0.64	0.86	0.95	0.98	0.98	0.99

Table 6: The percentage of upgrades (downgrades) of the actual and the predicted ratings and the percentage of correctly predicted upgrades (downgrades).

		Upgrades			**Downgrades**		
		Actual	Predicted	%Correct	Actual	Predicted	%Correct
S&P	Multi-year ordered probit	0.16	0.22	0.45	0.11	0.18	0.49
	Single-year ordered probit	0.16	0.27	0.45	0.11	0.23	0.45
	Pooled ordered probit	0.16	0.19	0.29	0.11	0.13	0.32
	SUR linear regression	0.16	0.19	0.34	0.11	0.16	0.46
	OLS linear regression	0.16	0.27	0.40	0.11	0.23	0.45
Moody's	Multi-year ordered probit	0.13	0.21	0.39	0.09	0.19	0.62
	Single-year ordered probit	0.13	0.28	0.43	0.09	0.24	0.52
	Pooled ordered probit	0.13	0.21	0.33	0.09	0.16	0.44
	SUR linear regression	0.13	0.17	0.43	0.09	0.16	0.59
	OLS linear regression	0.13	0.27	0.45	0.09	0.24	0.53
Fitch	Multi-year ordered probit	0.14	0.21	0.46	0.09	0.16	0.57
	Single-year ordered probit	0.14	0.25	0.41	0.09	0.22	0.59
	Pooled ordered probit	0.14	0.21	0.30	0.09	0.13	0.44
	SUR linear regression	0.14	0.17	0.40	0.09	0.15	0.59
	OLS linear regression	0.14	0.28	0.40	0.09	0.26	0.61

10% for downgrades), in line with the findings of Hu *et al.* (2002). This lower number of actual rating changes can be explained by a trade-off in the CRAs' rating system between stability and accuracy (Cantor and Mann, 2007; Gaillard, 2012).

5. Conclusion

This paper compares the importance of ten determinants of sovereign credit ratings over time for the three main credit rating agencies, using a sample of 90 countries for the years 2002-2015. Applying a composite marginal likelihood estimation approach, we estimate a multi-year ordered probit model.

We provide empirical evidence that the credit rating agencies changed their sovereign credit rating assessment after the start of the European debt crisis in 2009. The financial balance, the economic development and the external debt became substantially more important after 2009, and the effect of eurozone membership switched from positive to negative. In addition, GDP growth and government debt, as well as their interaction, gained much importance, such that the positive effect of GDP growth on the credit rating became considerable, especially for highly indebted sovereigns, and that the negative effect of government debt became large, especially for low growth countries. Very recent papers confirm some of our findings. Comparing estimated single-year linear regression coefficients between the year 2007 and the year 2015, Amstad and Packer (2015) find that the government debt to GDP ratio, the GDP growth rate and the flexibility of the exchange rate regime was more important for the latter year. Also Boumparis *et al.* (2015) and Giacomino (2013) find that government debt became more important after 2008, using a panel linear regression model.

We believe that our empirical model with ten determinants provides a good understanding of the credit rating process: the in sample predictions have an average absolute error of about 1.5 notches, they could correctly predict about 50% of rating up- and downgrades and about 55% of them lie within one notch of the actual rating. Still, we acknowledge that our model remains a simplified representation of the complex sovereign credit rating process of the CRAs, which incorporates hundreds of variables as well as subjective judgment, and which can vary the relevance of the different determinants across countries (S&P, 2008). A related limitation is that we model end of year sovereign credit ratings using variables, such as GDP per capita and inflation, which are published and revised several months after the end of the year. On the other hand, rating agencies do have access to other data series such as surveys, which could inform on the current values of these variables.

Finally, our approach of analyzing the ratio of each coefficient relative to the coefficient of GDP per capita, has the limitation that a change in this ratio does not inform per se on whether the importance of the numerator variable has changed or whether the importance of the denominator variable GDP per capita has changed. However, both a seemingly unrelated linear regression analysis and the linear regression model of Amstad and Packer (2015) indicate that the importance of GDP per capita was relatively constant over time (ignoring the bias that results from applying such linear regression models). This strengthens our interpretation that changes in the ratio are driven by changes in the importance of the numerator variable.

Our results provide insight in the sovereign credit rating process that are relevant to credit rating agencies, financial investors and governments. The model can be used by credit rating agencies as an empirical approximation for their credit rating process. Furthermore, predictions of the credit rating for non-rated countries can be obtained. Finally, our quantification of the determinants of sovereign credit ratings can help sovereigns to better understand the drivers of their credit rating.

Acknowledgments

We are thankful to Ferre De Graeve, Geert Dhaene and Frank Smets for useful discussions and helpful comments. Also, we are grateful to Peter Chovanec and Wouter Simons for their research assistance. Finally, financial support from the Agency for Innovation by Science and Technology (IWT) in Flanders is kindly acknowledged.

References

Afonso, A., Gomes, P., and Rother, P. (2007). What 'hides' behind sovereign debt ratings? Working Paper 711, European Central Bank.

Afonso, A., Gomes, P., and Rother, P. (2009). Ordered response models for sovereign debt ratings. *Applied Economics Letters*, **16**(8), 769–773.

Afonso, A., Gomes, P., and Rother, P. (2011). Short and long run determinants of sovereign debt credit ratings. *International Journal of Finance & Economics*, **16**(1), 1–15.

Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research*, **28**(2), 186–208.

Amstad, M. and Packer, F. (2015). Sovereign ratings of advanced and emerging economies after the crisis. *BIS Quarterly review*, pages 77–91.

Beers, D. T. and Nadeau, J.-S. (2015). Database of sovereign defaults. Technical Report 101, Bank of Canada.

Bhat, C., Varin, C., and Ferdous, N. (2010). A comparison of the maximum simulated likelihood and composite marginal likelihood estimation approaches in the context of the multivariate ordered response model. In W. Greene and R. Hill, editors, *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, volume 26, pages 65–106. Emerald Group Publishing Limited.

Bissoondoyal-Bheenick, E. (2005). An analysis of the determinants of sovereign ratings. *Global Finance Journal*, **15**(3), 251–280.

Bissoondoyal-Bheenick, E., Brooks, R., and Yip, A. Y. (2006). Determinants of sovereign ratings: A comparison of case-based reasoning and ordered probit approaches. *Global Finance Journal*, **17**(1), 136–154.

Boes, S. and Winkelmann, R. (2006). The effect of income on positive and negative subjective well-being. SOI - Working Papers 0605, Socioeconomic Institute - University of Zurich.

Boumparis, P., Milas, C., and Panagiotidis, T. (2015). Has the crisis affected the behavior of the rating agencies? Panel evidence from the eurozone. *Economics Letters*, **136**(C), 118–124.

Cantor, R. and Mann, C. (2007). Analyzing the tradeoff between ratings accuracy and stability. *Journal of Fixed Income*, pages 60–68.

Cantor, R. and Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Economic Policy Review*, pages 37–53.

Christensen, R. H. B. (2015). Analysis of ordinal data with cumulative link models. Estimation with the R package Ordinal.

De Grauwe, P. and Ji, Y. (2013). Self-fulfilling crises in the eurozone: An empirical test. *Journal of International Money and Finance*, **34**, 15 – 36.

Elkhoury, M. (2007). Credit rating agencies and their potential impact on developing countries. UNCTAD Discussion Papers 186, United Nations Conference on Trade and Development.

Ferdous, N., Eluru, N., Bhat, C. R., and Meloni, I. (2010). A multivariate ordered-response model system for adults' weekday activity episode generation by activity purpose and social context. *Transportation Research Part B: Methodological*, **44**(8-9), 922–943.

Fitch (2014). Sovereign rating criteria. Technical report.

Gaillard, N. (2012). *A century of sovereign ratings*. Springer-Verlag New York.

Gartner, M., Griesbach, B., and Jung, F. (2011). PIGS or lambs? the European sovereign debt crisis and the role of rating agencies. *International Advances in Economic Research*, **17**(3), 288–299.

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *Journal of Computational and Graphical Statistics*, **1**(2), 141–149.

Giacomino, P. (2013). Are sovereign credit ratings pro-cyclical? a controversial issue revisited in light of the current financial crisis. *Rivista di Politica Economica*, (4), 79–111.

Greene, W. and Hensher, D. (2010). *Modeling ordered choices: a primer*. Cambridge University Press, New York.

Hill, P., Brooks, R., and Faff, R. (2010). Variations in sovereign credit quality assessments across rating agencies. *Journal of Banking & Finance*, **34**(6), 1327–1343.

Hoetker, G. (2004). Confounded coefficients: Extending recent advances in the accurate comparison of logit and probit coefficients across groups. Working Paper 030100.

Hoetker, G. (2007). The use of logit and probit models in strategic management research: Critical issues. *Strategic Management Journal*, **28**(4), 331–343.

Hu, Y.-T., Kiesel, R., and Perraudin, W. (2002). The estimation of transition matrices for sovereign credit ratings. *Journal of Banking & Finance*, **26**(7), 1383–1406.

Kiff, J., Holland, A., Kisser, M., Nowak, S., Saab, S., Schumacher, L., van der Hoorn, H., and Westin, A.-M. (2010). The uses and abuses of sovereign credit ratings. In *Global Financial Stability Report*, volume 10/2010. International Monetary Fund.

McKelvey, R. D. and Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, **4**(1), 103–120.

Monfort, B. and Mulder, C. (2000). Using credit ratings for capital requirements on lending to emerging market economies - possible impact of a new Basel accord. Technical Report 00/69, IMF Working Papers.

Moody's (2015a). Moody's statistical handbook country credit. Technical report, Moody's Investor Service.

Moody's (2015b). Sovereign bond ratings. Technical report, Moody's Investor Service.

Mora, N. (2006). Sovereign credit ratings: guilty beyond reasonable doubt? *Journal of Banking & Finance*, **30**(7), 2041–2062.

S&P (2008). Sovereign credit ratings: A primer. Technical report, Standard and Poor's.

S&P (2014a). How S&P factors in monetary union membership in its sovereign ratings. Technical report, Standard and Poor's.

S&P (2014b). Sovereign rating methodology. Technical report, Standard and Poor's.

Train, K. (1998). Recreation demand models with taste differences over people. *Land economics*, **74**(2), 230–239.

Train, K. (2003). *Discrete choice methods with simulation*. Cambridge University Press.

Varin, C. and Czado, C. (2010). A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics*, **11**(1), 127–138.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, **21**(1), 5–42.

Zhao, Y. and Joe, H. (2005). Composite likelihood estimation in multivariate data analysis. *The Canadian Journal of Statistics*, **33**(3), 335–356.

Appendix A. Data Appendix

Table A.1: The sovereign credit rating categories used by S&P, Moody's and Fitch.

Category number	S&P	Moody's	Fitch
21	AAA	Aaa	AAA
20	AA+	Aa1	AA+
19	AA	Aa2	AA
18	AA-	Aa3	AA-
17	A+	A1	A+
16	A	A2	A
15	A-	A3	A-
14	BBB+	Baa1	BBB+
13	BBB	Baa2	BBB
12	BBB-	Baa3	BBB-
11	BB+	Ba1	BB+
10	BB	Ba2	BB
9	BB-	Ba3	BB-
8	B+	B1	B+
7	B	B2	B
6	B-	B3	B-
5	CCC+	Caa1	CCC+
4	CCC	Caa2	CCC
3	CCC-	Caa3	CCC-
2	CC	Ca	CC
1	SD/D	C	D/RD

Table A.2: The countries included in the sample for S&P, Moody's and Fitch, denoted by 'x'.

	S&P	Moody's	Fitch		S&P	Moody's	Fitch
Argentina	x	x	x	Latvia	x	x	x
Australia	x	x	x	Lebanon	x	x	x
Austria	x	x	x	Lithuania	x	x	x
Bahamas, The		x		Luxembourg	x	x	x
Bahrain	x	x	x	Malta	x	x	x
Barbados	x	x		Mauritius		x	
Belgium	x	x	x	Mexico	x	x	x
Belize	x	x		Morocco	x	x	
Bolivia	x	x		Netherlands	x	x	x
Botswana	x			New Zealand	x	x	x
Brazil	x	x	x	Nicaragua		x	
Bulgaria	x	x	x	Norway	x	x	x
Canada	x	x	x	Oman	x	x	
Chile	x	x	x	Pakistan	x	x	
China	x	x	x	Panama	x	x	x
Colombia	x	x	x	Papua New Guinea	x	x	
Costa Rica	x	x	x	Paraguay	x	x	
Croatia	x	x	x	Peru	x	x	x
Cyprus	x	x	x	Philippines	x	x	x
Czech Republic	x	x	x	Poland	x	x	x
Denmark	x	x	x	Portugal	x	x	x
Dominican Republic	x	x		Qatar	x	x	
Ecuador	x	x	x	Romania	x	x	x
Egypt, Arab Rep.	x	x	x	Russian Federation	x	x	x
El Salvador	x	x	x	Saudi Arabia		x	
Estonia	x	x	x	Senegal	x		
Fiji		x		Singapore	x	x	x
Finland	x	x	x	Slovak Republic	x	x	x
France	x	x	x	Slovenia	x	x	x
Germany	x	x	x	South Africa	x	x	x
Greece	x	x	x	Spain	x	x	x
Guatemala	x	x		Suriname	x		
Honduras		x		Sweden	x	x	x
Hong Kong SAR, China	x	x	x	Switzerland	x	x	x
Hungary	x	x	x	Taiwan	x	x	x
Iceland	x	x	x	Thailand	x	x	x
India	x	x	x	Trinidad and Tobago	x	x	
Indonesia	x	x	x	Tunisia		x	x
Ireland	x	x	x	Turkey	x	x	x
Israel	x	x	x	Ukraine	x	x	x
Italy	x	x	x	United Arab Emirates		x	
Jamaica	x	x		United Kingdom	x	x	x
Japan	x	x	x	United States	x	x	x
Jordan	x	x		Uruguay	x	x	x
Kazakhstan	x	x	x	Venezuela, RB	x	x	x
Korea, Rep.	x	x	x	Vietnam	x	x	x
Kuwait	x	x	x				

Appendix B. The composite likelihood estimator of the multi-year ordered probit model

Appendix B.1. The composite likelihood function

The logarithm of the composite likelihood function $L^C(\theta)$, defined in (5), can be written as

$$\log L^C(\theta) = \sum_{i=1}^{N}\sum_{s=1}^{T-1}\sum_{t=s+1}^{T}\sum_{j=1}^{C_s}\sum_{k=1}^{C_t} I[y_{is} = j, y_{it} = k] \times \log P\left(Y_{is} = j, Y_{it} = k\right), \tag{B.1}$$

where y_{is} and y_{it} denote the *observed* category of variables Y_{is} and Y_{it}. $P\left(Y_{is} = j, Y_{it} = k\right)$ is given by

$$P\left(Y_{is} = j, Y_{it} = k\right) = P\left(\tau_{is}^{j-1} < \nu_{is} < \tau_{is}^{j}, \tau_{it}^{k-1} < \nu_{it} < \tau_{it}^{k}\right)$$
$$= \Phi_2(\tau_{is}^{j}, \tau_{it}^{k}; \Sigma_{st}) + \Phi_2(\tau_{is}^{j-1}, \tau_{it}^{k-1}; \Sigma_{st}) - \Phi_2(\tau_{is}^{j}, \tau_{it}^{k-1}; \Sigma_{st}) - \Phi_2(\tau_{is}^{j-1}, \tau_{it}^{k}; \Sigma_{st}), \tag{B.2}$$

where $\Phi_2(\cdot, \cdot; \rho)$ is the cdf of the bivariate normal distribution function with correlation parameter ρ and unit variances, and where τ_{it}^{l} is defined as

$$\tau_{it}^{l} = \tau_{t}^{l} - \beta_t x_{it},$$

for i in $1, ..., N$, t in $1, ..., T$ and l in $0, ..., C_t$.

Appendix B.2. The covariance matrix of the composite likelihood estimator

The covariance matrix of the composite likelihood estimator $Cov(\hat{\theta})$ equals the inverse of the Godambe's sandwich information matrix $G(\theta)$ (Zhao and Joe, 2005)

$$Cov(\hat{\theta}) = G(\theta)^{-1} = H(\theta)^{-1} J(\theta) H(\theta)^{-1}, \tag{B.3}$$

where

$$J(\theta) = E\left[\left(\frac{\partial \log L^C(\theta)}{\partial \theta}\right)\left(\frac{\partial \log L^C(\theta)}{\partial \theta}\right)'\right]$$
$$H(\theta) = E\left[\frac{\partial^2 \log L^C(\theta)}{\partial \theta \partial \theta'}\right],$$

where θ is the vector collecting all unknown elements, as defined in Section 3.1. The matrices $H(\theta)$ and $J(\theta)$ can be estimated as follows (Bhat *et al.*, 2010; Ferdous *et al.*, 2010; Varin *et al.*, 2011)

$$\hat{J}(\hat{\theta}) = \sum_{i=1}^{N} \left[\left(\frac{\partial \log L_i^C(\theta)}{\partial \theta} \right) \left(\frac{\partial \log L_i^C(\theta)}{\partial \theta} \right)' \right]_{\hat{\theta}}$$

$$= \sum_{i=1}^{N} \left(\sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \sum_{j=1}^{C_s} \sum_{k=1}^{C_t} \frac{I[y_{is}=j, y_{it}=k]}{P(Y_{is}=j, Y_{it}=k)} \frac{\partial P\left(Y_{is}=j, Y_{it}=k\right)}{\partial \theta} \right)_{\hat{\theta}}$$

$$\times \left(\sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \sum_{j=1}^{C_s} \sum_{k=1}^{C_t} \frac{I[y_{is}=j, y_{it}=k]}{P(Y_{is}=j, Y_{it}=k)} \frac{\partial P\left(Y_{is}=j, Y_{it}=k\right)}{\partial \theta} \right)'_{\hat{\theta}} \tag{B.4}$$

and

$$\hat{H}(\hat{\theta}) = \sum_{i=1}^{N} \left[\frac{\partial^2 \log L_i^C(\theta)}{\partial \theta \partial \theta'} \right]_{\hat{\theta}}$$

$$= \sum_{i=1}^{N} \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \sum_{j=1}^{C_s} \sum_{k=1}^{C_t} I[y_{is}=j, y_{it}=k] \left[\frac{\partial^2 \log P\left(Y_{is}=j, Y_{it}=k\right)}{\partial \theta \partial \theta'} \right]_{\hat{\theta}}$$

$$= - \sum_{i=1}^{N} \sum_{s=1}^{T-1} \sum_{t=s+1}^{T} \sum_{j=1}^{C_s} \sum_{k=1}^{C_t} \left[\frac{I[y_{is}=j, y_{it}=k]}{P(Y_{is}=j, Y_{it}=k)^2} \right.$$

$$\left. \frac{\partial P\left(Y_{is}=j, Y_{it}=k\right)}{\partial \theta} \frac{\partial P\left(Y_{is}=j, Y_{it}=k\right)}{\partial \theta}' \right]_{\hat{\theta}}, \tag{B.5}$$

where L_i^C is defined in (6) and the $\hat{\theta}$ subscript denotes that the function is evaluated at the composite likelihood estimator $\hat{\theta}$.

For $1 \leq i \leq N$, $1 \leq s < t \leq T$, j in $1, ..., C_s$ and k in $1, ..., C_t$, the nonzero elements of the vector $\frac{\partial P(Y_{is}=j, Y_{it}=k)}{\partial \theta}$ used in equations (B.4) and (B.5) can be computed from (B.2) and are given below:

- the component corresponding to τ_s^{j-1} with $2 \leq j \leq C_s$:

$$\phi(\tau_{is}^{j-1}) \left(\Phi(\frac{\tau_{it}^{k-1} - \Sigma_{st} \tau_{is}^{j-1}}{\sqrt{1-\Sigma_{st}^2}}) - \Phi(\frac{\tau_{it}^{k} - \Sigma_{st} \tau_{is}^{j-1}}{\sqrt{1-\Sigma_{st}^2}}) \right), \tag{B.6}$$

- the component corresponding to τ_s^{j} with $1 \leq j \leq C_s - 1$:

$$\phi(\tau_{is}^{j}) \left(\Phi(\frac{\tau_{it}^{k} - \Sigma_{st} \tau_{is}^{j}}{\sqrt{1-\Sigma_{st}^2}}) - \Phi(\frac{\tau_{it}^{k-1} - \Sigma_{st} \tau_{is}^{j}}{\sqrt{1-\Sigma_{st}^2}}) \right), \tag{B.7}$$

- the component corresponding to τ_t^{k-1} with $2 \leq k \leq C_t$:

$$\phi(\tau_{it}^{k-1}) \left(\Phi(\frac{\tau_{is}^{j-1} - \Sigma_{st} \tau_{it}^{k-1}}{\sqrt{1-\Sigma_{st}^2}}) - \Phi(\frac{\tau_{is}^{j} - \Sigma_{st} \tau_{it}^{k-1}}{\sqrt{1-\Sigma_{st}^2}}) \right), \tag{B.8}$$

- the component corresponding to τ_t^k with $1 \leq k \leq C_t - 1$:

$$\phi(\tau_{it}^k)\left(\Phi(\frac{\tau_{is}^j - \Sigma_{st}\tau_{it}^k}{\sqrt{1-\Sigma_{st}^2}}) - \Phi(\frac{\tau_{is}^{j-1} - \Sigma_{st}\tau_{it}^k}{\sqrt{1-\Sigma_{st}^2}})\right), \tag{B.9}$$

- the p components corresponding to β_s:[10]

$$(-x_{is})\Big\{(\phi(\tau_{is}^j)\Phi(\frac{\tau_{it}^k - \Sigma_{st}\tau_{is}^j}{\sqrt{1-\Sigma_{st}^2}}) + \phi(\tau_{is}^{j-1})\Phi(\frac{\tau_{it}^{k-1} - \Sigma_{st}\tau_{is}^{j-1}}{\sqrt{1-\Sigma_{st}^2}})$$
$$-\phi(\tau_{is}^j)\Phi(\frac{\tau_{it}^{k-1} - \Sigma_{st}\tau_{is}^j}{\sqrt{1-\Sigma_{st}^2}}) - \phi(\tau_{is}^{j-1})\Phi(\frac{\tau_{it}^k - \Sigma_{st}\tau_{is}^{j-1}}{\sqrt{1-\Sigma_{st}^2}})\Big\}, \tag{B.10}$$

- the p components corresponding to β_t:

$$(-x_{it})\Big\{\phi(\tau_{it}^k)\Phi(\frac{\tau_{is}^j - \Sigma_{st}\tau_{it}^k}{\sqrt{1-\Sigma_{st}^2}}) + \phi(\tau_{it}^{k-1})\Phi(\frac{\tau_{is}^{j-1} - \Sigma_{st}\tau_{it}^{k-1}}{\sqrt{1-\Sigma_{st}^2}})$$
$$-\phi(\tau_{it}^k)\Phi(\frac{\tau_{is}^{j-1} - \Sigma_{st}\tau_{it}^k}{\sqrt{1-\Sigma_{st}^2}}) - \phi(\tau_{it}^{k-1})\Phi(\frac{\tau_{is}^j - \Sigma_{st}\tau_{it}^{k-1}}{\sqrt{1-\Sigma_{st}^2}})\Big\}, \tag{B.11}$$

- the component corresponding to ρ:

$$|t-s|\rho^{|t-s|-1}$$
$$\times\left(\phi_2(\tau_{is}^j, \tau_{it}^k; \Sigma_{st}) + \phi_2(\tau_{is}^{j-1}, \tau_{it}^{k-1}; \Sigma_{st}) - \phi_2(\tau_{is}^j, \tau_{it}^{k-1}; \Sigma_{st}) - \phi_2(\tau_{is}^{j-1}, \tau_{it}^k; \Sigma_{st})\right), \tag{B.12}$$

where $\Phi(\cdot)$ denotes the standard normal distribution function, $\phi(\cdot)$ denotes the standard normal density function and $\phi_2(\cdot, \cdot; \Sigma_{st})$ denotes the bivariate normal density function with correlation parameter Σ_{st} and unit variances.

Appendix B.3. Implementation of the composite likelihood estimator

We perform two reparameterizations. First, we write the autoregressive parameter ρ, between -1 and 1, as the hyperbolic tangent transformation of an unrestricted parameter ρ_{atanh}. Second, in line with Greene and Hensher (2010), we reparametrize the threshold coefficients τ_t^l to ensure that the ordering $\tau_t^i < \tau_t^j$ for $i < j$ is preserved. Define γ_t^j, for each t in $1, ..., T$ as

$$\tau_t^1 = \gamma_t^1$$
$$\tau_t^j = \tau_t^{j-1} + \exp(\gamma_t^j) \quad \text{for } j \text{ in } 2, ..., C_t - 1.$$

[10]In (B.10), we use the convention that the first component equals zero when both $j = C_s$ and $k = C_t$ and that the second component equals zero when both $j = 1$ and $k = 1$. A similar convention applies for (B.11).

We maximize the composite likelihood using the BFGS algorithm implemented in the 'optim' function of the R package 'stats'. The gradient of the composite likelihood function, which is used in the BFGS optimization algorithm, is computed analytically from (B.1). The bivariate normal probabilities of the pairwise composite loglikelihood function in (B.1) are computed using the Genz (1992) algorithm implemented in the R package *mnormt*. The starting values for the parameters β_t and γ_t^j are chosen as the maximum likelihood estimates from the single-year ordered probit model. The starting values of the ρ_{atanh} parameter is the inverse hyperbolic tangent transformation of the average of the estimates of the off-diagonal elements of the estimated covariance matrix of the seemingly unrelated linear regression model to the power $1/|t-s|$, where s and t denote the row and column number.

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be